Data 698

Spring 2022

Jimmy Ng

2022-02-15

**Final Project Proposal (version 2)**

**Research Objective**

This research examines the efficacy of technical indicators and predictive modeling in optimizing a set of investment strategy for the S&P 500. Specifically, this research makes comparison of the S&P 500 2-year prior to and 2-year after the beginning of the global pandemic.

Below analogy drawing from a hypothetical experiment can illustrate the goal and design of this research study.

A/B testing: Imagine we are conducting A/B testing for treating a disease (or enhancing physical performance). We have 500 patients. We have two treatment plans, i.e., RSI, MACD. This is a longitudinal study, and we have two study trials. First, we follow each patient for 2-year prior to the start of pandemic (Y2018, Y2019). Second, we follow them for 2-year after the pandemic began (Y2020, Y2021). At the end of each trial, we measure their performance (return of investment, ROI) based on the treatment plans. In addition, we can build classification model based on features generating from these treatment plans. We will evaluate the treatment plans by drawing conclusion from the data (based on ROI).

**Hypotheses**

1) Null hypothesis (H01): all indicators make no significant difference in terms of return of investment (ROI), i.e., ROI is not significantly different from 0

2) Null hypothesis (H02): all indicators are statistically the same in terms of ROI

3) Null hypothesis (H03): there is no significant difference in terms of ROI (generated by the indicators) among sectors

4) Null hypothesis (H04): there is no significant difference in terms of ROI (generated by the indicators) before and after the pandemic

5) Null hypothesis (H05): a predictive model built from using the features generated by the technical indicators do not perform significantly better than traditional technical analysis, strategy

**Methodology**

We can conduct this hypothetical study using simulation on historical data of the S&P 500. In the first trial, we begin the simulation of investing $10,000 in January 2018 for each stock. The trading decision (“entry”, “exit”) is based on the indicators (and different set of parameters). We will measure the ROI by the end of December 2019. Similarly for the second trial, we start investing $10,000 in January 2020 until December 2021, and then we will measure the ROI at the end.

We will evaluate the efficacy of these technical indicators based on the performance of these 500 “subjects” in these two trials.  
  
**Sample Data**

This is open-source data, e.g., quantmod. Below is a mock-up of a data frame for the experimental simulation.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **symbol** | **company\_name** | **sector** | **trial** | **RSI** | **MACD** |
| A | Agilent Technologies Inc | Health Care | pre-pandemic | $3,000 | $2,000 |
| AAL | American Airlines Group | Industrials | pre-pandemic | $1,000 | $3,300 |
| AAP | Advance Auto Parts | Consumer Discretionary | pre-pandemic | ($500) | ($700) |
| AAPL | Apple Inc. | Information Technology | pre-pandemic | $2,500 | $2,000 |
| ABBV | AbbVie Inc. | Health Care | pre-pandemic | ($900) | $200 |